The Evolution of School Failure Risk During the 2000 Decade in Spain: Analysis of PISA Results with a Two-Level Logistic Model

Evolución del riesgo de fracaso escolar en España durante la década del 2000: Análisis de los resultados de PISA con un modelo logístico de dos niveles

Juan Manuel Guio Jaimes
Universidad de Barcelona
jguiojai7@alumnes.ub.edu

Álvaro Choi de Mendizábal
Universidad de Barcelona
alvarochoi@ub.edu

Abstract: School failure has been one of the principal problems of the Spanish educational system during recent decades. This paper offers a perspective of the evolution of the factors that have had a significant influence over the risk of school failure considering personal, household and school characteristics through multilevel logistic regression analyses of PISA 2000, 2003, 2006 and 2009 microdata.

Keywords: school failure; multilevel logistic regression; PISA; compulsory education.

Resumen: El fracaso escolar ha sido uno de los principales problemas del sistema educativo español durante las últimas décadas. Este trabajo ofrece una perspectiva de la evolución en los factores que tienen una influencia significativa sobre el riesgo de fracaso escolar considerando características personales, familiares y de las escuelas por medio de un análisis de regresión con un modelo logístico multinivel para los micro datos de PISA 2000, 2003, 2006 y 2009.

Palabras clave: fracaso escolar; regresión logística multinivel; PISA; educación secundaria obligatoria.

INTRODUCTION

Overcoming high rates of school failure is one of the most complex challenges faced by Spanish society. National and international studies, such as Fernández Enguita et al. (2010), and OECD reports (2010, 2011), address the problem of the low performance of Spanish students compared to their European peers. Results are clear: Spain faces higher levels of school failure and early school dropout.

The broad definition of school failure includes all forms of not achieving the educational objectives determined by society as the minimum necessary to be integrated into the labor market and to become a productive member of the community. Accordingly, the definition of school failure chosen in this paper includes all the individuals who are not able to complete compulsory secondary education (ESO) at the age of 16.

School failure in Spain appears to have structural characteristics, as it has been present in the educational system for more than 30 years, with figures fluctuating around 30% during the last two decades. Failure to accomplish the objective proposed by the Lisbon Strategy in 2010 and difficulties in achieving the Europe 2020 strategy of reducing the early school leaver rate in EU countries to less than 10% are indicators of issues in addressing the causes of the problem.

The aim of this paper is to analyze the evolution of the factors that determined school failure risk during the 2000 decade in Spain. In this paper, following the work of authors such as Schleicher (2007), school failure risk is defined as the probability of obtaining a score below level-2 in reading competency in the Programme for International Student Assessment (PISA). The selection of reading competency as the main area in this analysis is due to the emphasis that this competency has in two of the four PISA tests (2000 and 2009).

The analysis is performed using 2000, 2003, 2006 and 2009 PISA micro-data for Spain. This should allow the observation of variations in the determinants over time, and their importance as predictors of school failure risk, broadening the scope of previous works such as Calero et al. (2010) or Choi and Calero (in press) and permitting the introduction of methodological improvements.

Table 1 compares real school failure rates in Spain and the risk of school failure in all the PISA test competencies. The measure of school failure risk in PISA tends to underestimate the real volume of students who fail.

Table 1: Risk of school failure by competencies in the Spain PISA tests

	2000	2003	2006	2009
School failure (real)	26,6	28,7	30,8	25,9
Reading competence	16,7	18,5	19,1	20,0
Mathematics competence	25,0	20,4	17,2	21,8
Scientific Knowledge competence	20,6	20,3	14,0	17,7

Source: OECD-PISA, 2000, 2003, 2006 and 2009 databases and Ministerio de Educación, Cultura y Deporte (2011).

This paper is structured as follows: Section 2 explains and justifies the individual, household and school-level variables considered in the analysis. Section 3 presents the data and the methodology, while Section 4 discusses the results. Finally the main conclusions are presented in the last section.

Determinants of academic performance and school failure

The present section has been divided into three sub-sections according to the three blocks of explicative variables considered in this paper: personal, family and school characteristics.

Personal characteristics

The differences in the academic performance of students depend on a number of characteristics that are distinctive at an individual level, and have a direct influence on the probability of school success. Gender appears to be an important personal determinant that affects the academic performance of students. For example, there is a significant difference in the ESO graduation rates between males and females. Using year 2006-2007 data, Fernández Enguita (2010) found a variation of almost 14 percentage points for females over males in graduation rates. A capacity for organization, discipline and attention appears to be a more common characteristic of female students, who also seem to have an advantage in the learning processes. It is therefore to be expected that a positive relation will be found between obtaining a high score in the PISA reading test and being female. It is also anticipated, that a lower proportion of female students will be at risk of school failure compared with

their male peers. It is important to mention that this result does not hold for the PISA math test (Calero et al., 2010). Carrington et al. (2008) attempted to analyze whether the gender gap in performance could be explained by teacher gender, in a study of 11-year-old British students. However they found no evidence of differences in the attitudes or approaches of teachers related to the gender of their students.

Another characteristic linked to the degree of maturity of the pupils is the month of birth of the student. Crawford et al. (2011) found evidence that students born between January and March seem to obtain consistently better grades than students who were born from October to December. Similarly, García Montalvo (2011) in a recent study in Spain, using the TIMSS and PIRLS databases, found evidence of a positive relationship between being born in the first and second quarters of the year and educational achievement measured by the scores of the students in the PISA international standardized tests. Data provided by the Instituto Vasco de Investigación y Evaluación Educativa (Basque Institute for Educational Research and Assessment) (2009) point in the same direction as the academic results of 4º grade primary school and 2º grade ESO students are influenced by the quarter of the year in which the student was born, although these differences seem to dissipate as the students get older. Consequently, a variable that accounts for the month of birth has been introduced to test for differences in the possible outcomes of students in the PISA tests.

Finally, students who are significantly older than their own school cohort seem to have some disadvantages in motivation, engagement and performance compared to their peers in the appropriate grade for their age (Martin, 2009). There are significant examples in the literature that provide evidence of a negative relation between grade retention and educational outcomes (Holmes, 1989; Jimmerson, 2001). However this variable has been intentionally excluded from our estimations due to endogeneity issues between academic achievement and grade repetition[1]. Table 2 shows the proportion of students who obtain a low result in the PISA reading competency in the four PISA evaluations.

1 In preliminary estimations of the results presented in Section 4 the variable "grade retention" was included, the results suggesting a strong link between grade retention and academic performance. However, we chose not to incorporate the variable in the final estimations in order to avoid bias generated by the introduction of an endogenous variable.

Table 2: Probabilities of obtaining a grade under level-2 in the PISA reading test according to personal characteristics

PERSONAL VARIABLES	2000	2003	2006	2009
GRADE				
2º ESO	74,19%	68,06%	69,35%	66,96%
3º ESO	42,05%	39,32%	38,37%	35,78%
4º ESO	5,46%	9,68%	7,90%	7,38%
GENDER				
Male	20,81%	24,90%	25,14%	24,19%
Female	10,94	11,83%	13,00%	13,89%
BIRTH MONTH				
January to March	14,68%	17,59%	16,47%	16,87%
October to December	19,13%	21,17%	22,01%	20,43%
COUNTRY OF BIRTH				
Spain	15,34%	17,39%	17,81%	16,98%
Others	29,95%	34,34%	36,82%	35,65%
ORIGIN OF THE STUDENT				
National students	16,31%	17,85%	18,22%	18,43%
First generation immigrants	33,33%	40,27%	40,86%	38,23%
Second generation immigrants		29,54%	32,99%	35,45%
AGGREGATE MEAN	17,37%	19,95%	20,44%	19,98%

Source: Own compilation with OECD-PISA, 2000, 2003, 2006 and 2009 databases.

Not all the students in the sample were born in Spain. An increasing number of students among the four PISA waves used in this article have diverse nationalities and origins. Being an immigrant is an important characteristic that seems to increase the risk of school failure (Table 2) and is related with adaptation issues, such as language and cultural differences (McCarthy, 1998). A dummy variable that distinguishes those students born in Spain from those born outside the country is introduced in the 2000 and 2003 waves. In the 2006 and 2009 waves, two dummy variables are introduced to measure the effect of being either a first or second-generation immigrant on the risk of school failure, compared to native students. The language spoken at home appears to be an important determinant of the process of adaptation of immigrants to their new country (Entorf and Minoiu, 2005).

Household characteristics

Household attributes and material resources are two important aspects to be considered in the analysis of academic performance. The Coleman Report (1966) provided evidence that family background is the main factor in student academic performance.

Hanushek (1997) showed that differences in household environments, such as students living in single-parent families or coming from lower socio-economical backgrounds, are relevant for individual academic achievement. Haveman and Wolfe (1995) state that it is a household's background characteristics that have the most powerful effect on the academic achievement of students.

There is an important difference in the academic performance of students whose parents belong to a managerial or professional category compared to those from families where the parents are manual workers (Cohen, 1987). The former are a small segment of the population and have significant advantages in school achievement, grades and completion rates compared to the latter. Consequently the model presented in Section 3 includes a variable that describes the household member in the highest socio-economic category.

There is ample literature that shows the relationship between the educational level of the parents and the performance of students. Ferguson et al. (1996) posited that parental education accounted for about 24 percent of the variance in student's test scores; Reynolds and Temple (1998) affirm that the level of education of the parents is positively associated with test scores and negatively with grade retention. Consequently it is to be expected that an inverse relation will be observed between the level of education attained by parents and the probability of students being at risk of school failure.

Table 3 illustrates the percentage of students at risk of school failure according to the evaluation results in the PISA reading competency and considering family attributes.

Table 3: Probability of obtaining a grade under level-2 in the PISA reading test considering household characteristics

SCHOOL VARIABLES	2000	2003	2006	2009
OCCUPATION OF THE PARENTS				
Working	15,59%	16,77%	15,86%	15,32%
Not working	19,12%	22,03%	24,15%	22,37%

SCHOOL VARIABLES	2000	2003	2006	2009
Socio-economical category				
Skilled white-collar worker	2,91%	6,54%	8,36%	6,37%
Unskilled white-collar worker	8,03%	11,15%	11,87%	9,56%
Blue-collar worker	19,11%	20,78%	22,31%	22,05%
PARENTAL EDUCATION LEVEL				
Tertiary education	8,59%	10,51%	12,43%	10,84%
Compulsory secondary education (ESO)	11,17%	16,87%	16,74%	18,43%
Primary education	24,91%	25,42%	34,89%	30,47%
Did not finish primary education	45,26%	40,95%	46,33%	55,51%
HOME EDUCATIONAL RESOURCES				
Computer, calculator, books and dictionary	12,23%	14,80%	14,06%	13,56%
Does not have these resources	22,53%	27,01%	23,63%	22,04%
HOUSEHOLD CULTURAL POSSESSIONS				
Literature, poetry and works of art	10,71%	12,37%	13,88%	13,37%
Does not have these possessions	27,37%	26,50%	29,05%	29,55%
AGGREGATE MEAN	17,37%	19,95%	20,44%	19,98%

Source: Own compilation with OECD-PISA, 2000, 2003, 2006 and 2009 databases.

The possession of cultural resources is expected to have a negative relation to the risk of school failure (Berger et al., 2005). Similarly the possession of a large number of books is correlated with early reading competence in individuals (Aikens and Barbarin, 2008). Two variables that account for the household's cultural environment and the specific material possessions within the household are incorporated in the model, and it is predicted that they will reduce the probability of obtaining a score below level-2 in the PISA reading test.

School characteristics

School-level determinants refer to the characteristics of the schools, the type of students who enroll in them and their material resources and their allocation. The most relevant factors affecting the risk of school failure seem to be to a significant extent already set before the students enter school, but it is important to determine if the school magnifies or reduces the differences between students with diverse characteristics and risk factors.

One significant determinant associated with the characteristics of educational institutions is school ownership. State schools contain a larger proportion of immigrant students and they have students with a wider range of characteristics and family backgrounds. The graduation rate from ESO in private schools is almost 20 percent above the rate in public institutions (Ministerio de Educación, 2009).

Interaction between students sharing certain characteristics contributes to the enhancement or the reduction of the academic performance of peers (Coleman et al., 1966). Accordingly the following determinants are introduced: the proportion of females in the school population, the socio-economic characteristics of the students and also the educational level attained by parents.

Another important characteristic related to the interaction between students is their origin. We have therefore introduced a variable that measures the proportion of immigrants in the schools. Two different thresholds are used: 20% for the years 2000 and 2003, and 30% for the years 2006 and 2009. The use of two different thresholds is justified by the progressive increase in the percentage of immigrant students who entered the Spanish educational system during the 2000 decade.

Calero and Escardíbul (2007), using PISA-2003, found that there is no significant difference in the performance of the students at different types of schools in Spain, and Cordero et al. (2011) reach the same conclusion, demonstrating that this variable was not relevant in the academic achievement of the students tested in PISA-2009. Considering this empirical evidence, it is to be expected that differences in the ownership of the school are not relevant determinants of school failure risk if socio-economical characteristics are controlled for.

The controversial debate over the importance of a reduced teacher-student ratio in the classroom is also considered in the analysis. While authors such as Krueger (2002) argued for the need to reduce class size in order to improve the quality of education, Hanushek (2003) and Chingos (2010) consider student-teacher ratio to be a factor that does not enhance the quality of education or academic results. To test this, we include a variable that represents student-teacher ratio.

Even though the real effect of school material resources on the academic performance of students is a matter of debate due to the results of international (Hanushek, 2003) and national studies (Calero, Choi and Waisgrais, 2010), variables such as school size and the students-per-computer ratio have been included with the purpose of observing their evolution throughout the 2000 decade. However, as Lavy (2012) warns, resource analysis may not be adequate if endogeneity is not addressed, that is if the fact that schools with certain profiles have higher student-teacher ratios than others is not taken into consideration. This is clearly a research area to be explored in Spain in the future.

Finally, another group of variables included in the analysis covers the participation of schools in school budget allocation and course content. According to the results observed by Calero and Waisgrais (2009), the effect of these variables is not particularly significant. However, using data from different years, it is still possible to observe interesting results regarding the effect of the autonomy of schools on school failure risk.

Table 4 shows the rate of students who could not achieve results equal to or above level-2 in the PISA reading competency, according to school attributes.

Table 4: Probability of obtaining a grade under level-2 in the PISA reading test considering school characteristics

SCHOOL VARIABLES	2000	2003	2006	2009
SIZE OF THE COMMUNITY WHERE SCHOOLS ARE LOCATED				
Community +100.000 inhabitants	12,63%	17,18%	15,29%	13,55%
Community +1.000.000 inhabitants	10,64%	18,02%	17,45%	18,14%
TYPE OF SCHOOL				
Private government independent	4,68%	11,13%	10,14%	8,13%
Private government dependent	10,32%	13,02%	13,69%	11,69%
Public	21,06%	22,93%	23,74%	23,97%
SCHOOL SIZE (NUMBER OF STUDENTS)				
Over the average	13,93%	12,44%	14,29%	17,00%
Under the average	19,63%	20,29%	22,22%	20,66%
PERCENTAGE OF GIRLS				
Over the average	17,23%	17,27%	18,39%	18,61%
Under the average	14,39%	17,57%	19,86%	19,78%
RATIO STUDENTS-COMPUTERS				
Over the average	17,88%	22,34%	22,60%	21,28%
Under the average	14,76%	14,68%	16,91%	17,33%
RATIO STUDENTS-TEACHER				
Over the average	8,91%	10,65%	13,81%	12,71%
Under the average	20,54%	22,93%	23,25%	23,00%
RATIO IMMIGRANTS-NATIONAL				
Immigrant students over 20%	33,33%	38,14%	30,89%	26,34%
Immigrant students over 30%	28,81%	53,42%	40,00%	30,05%
EDUCATIONAL SCHOOL ENVIRONMENT				
Tertiary education	7,19%	8,93%	10,07%	10,65%

SCHOOL VARIABLES	2000	2003	2006	2009
EDUCATIONAL SCHOOL ENVIRONMENT				
Compulsory secondary education (ESO)14,12%	19,23%	19,84%	22,59%	
Primary education	26,16%	32,07%	30,64%	58,59%
SOCIO-PROFESSIONAL SCHOOL ENVIRONMENT				
Skilled white-collar parents	0,00%	0,00%	3,50%	1,96%
Unskilled white-collar parents	4,51%	6,50%	8,86%	7,13%
Blue-collar parents	18,32%	20,69%	20,79%	20,45%
BUDGET MANAGEMENT AUTONOMY				
School resposibility	16,22%	11,33%	15,18%	9,76%
Not a school resposibility	25,00%	21,04%	22,63%	19,96%
COURSE CONTENT AUTONOMY				
School resposibility	15,89%	17,83%	19,40%	17,57%
Not a school resposibility	17,72%	18,11%	18,60%	19,35%
AGGREGATE MEAN	17,37%	19,95%	20,44%	19,98%

Source: Own compilation with OECD-PISA, 2000, 2003, 2006 and 2009 databases.

DATA AND METHODOLOGY

This section is divided into two parts. The first part describes the data provided by the PISA evaluations, the method employed to gather the data and the strategy necessary to handle this particular kind of database. The second part of the section outlines the econometric technique and model best suited to fit the PISA databases.

Data

The database for this paper corresponds to the OECD Programme for International Student Assessment (PISA) implemented in the late nineties as a strategy for the periodic international evaluation of the general competence of 15 year-old students. The sample in the present work covers the four PISA evaluations completed during the 2000 decade.

Table 5 describes the size of the sample of students and schools considered for each one of the four PISA evaluations in the present analysis.

Table 5: Sample size and target population of the Spain PISA tests from 2000 to 2009

	2000	2003	2006	2009
Total population of 15 years old	451,685	454,064	439,415	433,224
Students sample	6,214	10,791	19,604	25,887
Weighted student participants	399,055	344,372	381,686	387,054
Schools sample	185	383	686	889

Source: OECD-PISA, 2000, 2003, 2006 and 2009 databases.

In each PISA evaluation, emphasis is placed on one specific competency. These competencies were reading in the 2000 test, mathematics in 2003, scientific knowledge in 2006 and again reading in 2009.

The PISA test consists of three survey forms, one for the students, one for the parents and one for the head of the school. In Spain the parent questionnaire was omitted and, consequently, the main sources of information were the students and the schools. The student questionnaire is designed to gather information about specific competencies and also background information regarding the personal and household characteristics of the students. The school questionnaire collects information from the schools and makes it possible to match information between students and schools.

The PISA sampling is carried out in two stages. In the first stage, a sample of schools is randomly selected from a list of eligible schools. In the second stage a subsample of 35 students aged 15 years is randomly selected within each school. Student and school level sampling weights are provided to correct marginal deviation from the random probability process of selection. Consequently, sampling weights have been used throughout our analysis to avoid bias in population parameter estimates[2].

The educational outputs of the PISA evaluation are the scores in the standardized tests that give different values to the abilities of the students. Non-observable random factors that can affect the test result are controlled for through a set of plausible values. This characteristic of the PISA tests implies the need to incorporate the plausible values for each competency in the analysis. Performance in each PISA competency is measured through a set of 5 plausible values. These values cannot be individually interpreted as scores, but as a set they are accurate in

2 OECD (2009a) provides a thorough description of the use of sampling weights with PISA.

describing the performance of the population, as they contain a random error variance component. Estimations must therefore be performed five times per competency and then averaged. Standard errors are calculated following the same method, and the total variance is determined adding the measurement error and the sampling variances.

Missing values, present in all the PISA evaluations, require a particular approach that allows the true nature of the data to be left unaffected. The methodology that seems most appropriate is the Multiple Imputation strategy (Rubin, 1987), a procedure by which missing data are imputed several times to produce different complete data estimates of the parameters. The estimated parameters are combined to produce an overall estimate of the complete data parameters with minimal effects on the standard error. Multiple imputation by chained equations was performed using the Stata package. This software offers a more flexible method of dealing with missing values, compared to fully-parametric methods like maximum likelihood and Bayesian analysis.

Methodology

The analysis of PISA data requires multilevel modeling in order to account for the hierarchical structure of the data and a logit-type specification for the binary response dependent variable. We use a two-level formulation proposed by Raudenbush and Bryk (2002), the first level corresponds to data from the students clustered within schools, and the second level captures the influence of school factors.

Traditional techniques are not suited to accounting for the hierarchical and clustered structure of the data. Multilevel regression takes into account the nested distribution of the data within larger units of concentration, calculating a different equation for each level of aggregation. These models not only identify the relations of different variables within the same level but also the influence of variables from one level to another.

The first level of the multilevel regression corresponds to *i* students selected in the second stage of the PISA survey and, the second level, to *j* schools sampled in the previous stage of the survey. The logistic random intercept for a dichotomous dependent variable is modeled according to Raudenbush and Bryk (2002).

Two level random-intercept fixed-slope logistic regression model

LEVEL 1 MODEL

1) $$\eta_{ij} = \beta_{0j} + \sum_{k=1}^{n} \beta_{1j} X_{kij} + \xi_{ij}$$

2) $$p_{ij} = \frac{exp\,(\eta_{ij})}{1 + exp(\eta_{ij})}$$

$$y_{ij} = 1\ with\ probability\ p_{ij}$$

$$y_{ij} = 0\ with\ probability\ 1 - p_{ij}$$

3) $$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \beta_{0j} + \sum_{k=1}^{n} \beta_{1j} X_{kij} + \xi_{ij}$$

LEVEL 2 MODEL

4) $$\beta_{0j} = \gamma_{00} + \sum_{1} \gamma_{01} Z_{lj} + u_{0j}$$

5) $$\beta_{1j} = \gamma_{10}$$

$$u_{0j} \sim N(0, \tau_{00})$$

FULL MODEL

6) $$\log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \gamma_{00} + \gamma_{10} X_{kij} + \gamma_{01} Z_{lj} + u_{0j} X_{kij} + u_{0j} + \xi_{ij}$$

In the equations x_{kij} represents the student level covariates and Z_{1j} school level covariates. $\beta_{0j} .. \beta_{1j}$ represent regression coefficients. $u_{0j} \sim N\,(0, \tau_{00})$ are school specific random intercepts, uncorrelated across schools and uncorrelated with covariates. ξ_{ij} ~ logistic are student-specific residuals, uncorrelated across students and schools, uncorrelated with u_{0j} and with covariates. All multilevel estimations have been performed using HLM 6.25, which follows, for two-level models, the methodology suggested by Pfefferman et al (1998).

RESULTS

The regression results are shown in Table 6. Annex A describes the variance reduction analysis. The interpretation of the odd-ratios depends on the specification of the variables and the sign of the coefficients. When the variable has a positive coefficient, every 0.1 over 1.0 represents a 10% increase in the probability that the student scores below level-2 in reading competency. On the contrary, if the coefficient is negative, every 0.1 under 1 represents a 10% decrease in the probability of obtaining a grade under level-2.

In two particular cases, variables were replaced due to the differences in the sample size and the information available in the database for the four PISA evaluations.

In the personal characteristics, the variable in the 2000 and 2003 regressions (COB) that distinguished students born in Spain from those born outside the country, was replaced by two variables, FGIM (first generation immigrant) and SGIM (second generation immigrant), for 2006 and 2009, with the purpose of illustrating the evolution of immigrant students in the school system in Spain in greater detail.

In school characteristics, as has been explained, the chosen threshold for the ratio of immigrant students/total students in the school was increased from 20% (IRATIO20) for 2000 and 2003 to 30% (IRATIO30) for 2006 and 2009 in order to account for the progressive arrival of immigrant students in Spain and to capture significant effects.

Two mechanisms were used complementarily to check the correlations between independent variables with a correlation matrix and also to test the variance inflation factors (VIFs)[3]. Level 1 and level 2 variables were inspected separately.

Table 6 shows the coefficients of the two-level logistic regressions, the signs of which reflect the relation between the explanatory variables (personal, household and school characteristics) and the dependent variable (probability of obtaining a score below level-2 in reading competency), and the odd-ratios or likelihood ratios and the robust standard errors.

3 We consider that if any of the VIF values is larger than 4 there are multicollinearity problems associated with the variable.

Table 6: Multilevel logistic regressions fixed effects. Estimation for the probability of obtaining a score below level-2 in the reading competency PISA evaluations

	2000 HLM COEFF. RATIO	2000 HLM ODD	2003 HLM COEFF. RATIO	2003 HLM ODD	2006 HLM COEFF. RATIO	2006 HLM ODD	2009 HLM COEFF. RATIO	2009 HLM ODD
CONSTANT	-0.981 *** (0.360)	0.4	-0.973 ** (0.343)	0.4	-0.611 *** (0.162)	0.5	-1.375 *** (0.150)	0.3
SEX (gender)	0.688 *** (0.111)	2.0	0.844 *** (0.083)	2.3	0.710 *** (0.062)	2.0	0.727 *** (0.059)	2.1
M_BIRTH1(born in the 1st quarter)	-0.072 (0.140)	0.9	-0.045 (0.093)	1.0	-0.221 *** (0.081)	0.8	-0.141 ** (0.067)	0.9
M_BIRTH3 (born in the 4th quarter)	0.224 * (0.115)	1.3	0.072 (0.092)	1.1	0.160 * (0.075)	1.2	0.179 *** (0.070)	1.2
SEC1(skilled white-collar worker)	0.287 (0.314)	1.3	-0.138 (0.231)	0.9	-0.001 (0.174)	1.0	-0.004 (0.166)	1.0
SEC2(non-skilled white-collar)	-0.247 (0.178)	0.8	-0.226 (0.129)	0.8	-0.332 *** (0.100)	0.7	-0.378 *** (0.100)	0.7
COB (country of birth)	-0.412 (0.244)	0.7	-0.301 (0.215)	0.7				
FGIM (1st generation immigrant)					0.939 *** (0.148)	2.6	0.983 *** (0.098)	2.7
SGIM (2nd generation immigrant)					-0.086 (0.336)	0.9	0.307 (0.280)	1.4
OCCP (occupation of the parents)	0.243 * (0.108)	1.3	-0.030 (0.084)	1.0	-0.395 *** (0.078)	0.7	-0.086 (0.063)	0.9
HELP (Highest education level parents)	-0.054 *** (0.014)	0.9	-0.029 * (0.015)	1.0	-0.048 *** (0.011)	1.0	-0.045 *** (0.012)	1.0
HEDR (Home educational resources)	-0.443 *** (0.100)	0.6	-0.423 *** (0.084)	0.7	-0.387 *** (0.069)	0.7	-0.399 *** (0.072)	0.7
CULT (Cultural possessions/family)	-0.591 *** '(0.108)	0.6	-0.458 *** (0.085)	0.6	-0.559 *** (0.076)	0.6	-0.601 *** (0.064)	0.5
ST1 (Community +100.000 inhabit.)	0.023 (0.189)	1.0	0.155 (0.175)	1.2	-0.024 (0.133)	1.0	-0.113 (0.155)	0.9
ST2 (Community +1.000.000 inhab.)	-0.213 (0.273)	0.8	-0.255 (0.288)	0.8	-0.419 ** (0.215)	0.7	-0.513 ** (0.235)	0.6

	2000 HLM COEFF. RATIO	2000 HLM ODD	2003 HLM COEFF. RATIO	2003 HLM ODD	2006 HLM COEFF. RATIO	2006 HLM ODD	2009 HLM COEFF. RATIO	2009 HLM ODD
TOS1 (Type of school: Private)	-0.689 (0.436)	0.5	-0.114 (0.368)	0.9	-0.290 (0.273)	0.8	-0.006 (0.359)	1.0
TOS2 (T.of S. Private Government dependent)	-0.308 (0.328)	0.7	-0.148 (0.257)	0.9	-0.326 * (0.176)	0.7	-0.159 (0.260)	0.9
SCHLSIZE (Size of school by students)	-0.090 (0.204)	0.9	-0.267 (0.198)	0.8	-0.099 (0.125)	0.9	0.042 (0.137)	1.0
PCGIRLS (girls/school ratio)	-0.098 (0.180)	0.9	-0.279 ** (0.138)	0.8	0.010 (0.118)	1.0	-0.028 (0.110)	1.0
RATCOMP (student/ computer ratio)	-0.225 (0.166)	0.8	0.277 (0.204)	1.3	-0.046 (0.113)	1.0	0.189 (0.122)	1.2
STRATIO (student/ teacher ratio)	-0.170 (0.326)	0.8	-0.121 (0.305)	0.9	-0.073 (0.168)	0.9	-0.212 (0.227)	0.8
IRATIO20 (Immigrant +20% population)	0.836 *** (0.290)	2.3	0.917 *** (0.320)	2.5				
IRATIO30 (Immigrant +30% population)					0.370 * (0.191)	1.5	0.118 (0.181)	1.1
CLM3 (Most parents have tertiary education – school environment)	-0.449 * (0.254)	0.6	-0.384 (0.244)	0.7	-0.280 * (0.168)	0.8	-0.351 ** (0.153)	0.7
SPLP2 (Most parents are whitecollar category – school environment)	-0.368 (0.321)	0.7	-0.572 * (0.284)	0.6	-0.575 *** (0.246)	0.6	-0.169 (0.276)	0.8
B_MNGMENT (School budget autonomy)	-0.636 * (0.338)	0.5	-0.066 (0.304)	0.9	-0.166 (0.132)	0.8	-0.497 ** (0.234)	0.6
C_CONTENT (School curricular content autonomy)	-0.019 (0.196)	1.0	-0.072 (0.157)	0.9	0.074 (0.123)	1.1	-0.125 (0.171)	0.9
Observations	6,214		10,791		19,604		25,887	
Number of schools	185		383		686		889	

*** p<0.1 ** p<0.05, * p<0.01,

Note 1: *** p<0.01, ** p<0.05,* p<0.1, Standard errors in parentheses.
Note 2: The base category is constructed around these characteristics: Personal variables: Being female; being born between April and September; being born outside Spain. Household variables: Family whose parents belong to the blue-collar socio-professional category; parents not economically active; the household does not own educational resources and cultural possessions. School variables: Community has less than 100.000 inhabitants; state school; size of school by number of students, under or equal the average; percentage of girls in the school, under or equal the average; student-computer ratio, under or equal the average; student-teacher ratio, under or equal the average; immigrant students ratio, under or equal 20% of the population; academic environment of the schools consisting of parents who have attained primary and secondary level of education; the socio-professional environment of the schools, consisting of parents who belong to the blue-collar categories; school without autonomy in budget management and course content.

Personal variables

The first variable with considerable statistical significance in the regressions for the four years is the gender of the student. Being male increases one's chance of obtaining a result under level-2 in reading competency by over 100%. This result is consistent in all the regressions: 100% in 2000, 130% in 2003, 100% in 2006 and 110% in 2009.

This result implies that for every female student who obtains an unsatisfactory result in reading competency, there are at least two males in the same situation. The uniformity of the results over time seems to support the conclusions of Bertrand and Pan (2011) that showed that females have higher rates of success in the school system.

As suggested by Crawford et al. (2011), the month of birth seems to have a close relation to academic achievement and the cognitive skills of the students. This is not a surprising result as Sprietsma (2010), using PISA 2003 data, found a positive long term effect for relative age when entering primary school in 10 out of 16 countries analyzed.

The negative sign of the coefficient for the variable for students who were born from January to March, suggests that students born in the first quarter of the year are less likely to obtain a deficient result in the reading competency of the PISA evaluation. The same pattern is observed for the variable that describes the students born in the last quarter of the year, but with the opposite sign, indicating that students born from October to December are more prone to fail to obtain a result over level-2 in reading competency. These results also seem to support the findings by Robertson (2011), which show that students born in the first quarter of the year have a clear advantage in math and reading tests.

Regarding the origin of the students, a single variable that represents students who were born in Spain and whose parents are not from outside Spain was intro-

duced for the 2000 and 2003 PISA evaluations. Not being an immigrant reduced the probability of being at risk of failure in reading competency by 30%, compared with first and second generation immigrant students.

For 2006 and 2009, two variables were incorporated. The first variable corresponds to first generation immigrants and, as was anticipated, this variable was statistically significant for both years. The odd-ratios suggest that students born outside Spain and whose parents are first generation immigrants had a 160% higher risk of obtaining poor results in the reading test in 2006 and 170% in 2009, compared to their peers born in Spain.

The second variable corresponds to second generation immigrant students, those born in Spain but whose parents are both from outside Spain. The regression indicates that the variable is statistically non-significant for 2006 and 2009. These results suggest that second generation immigrant students do not perform differently to students born in Spain whose parents are also born within the country. Second generation immigrants obtain remarkably better results in the reading tests compared to first generation immigrant students.

Household variables

The variables describing family characteristics are divided into two categories. The first, attempts to characterize the family within a specific socio-economic and professional group, considering the educational achievement of the parents. The second group of variables represents the underlying relationship between achievement at school and the educational material and cultural resources of the household.

Socio-economic composition and educational level of the household

The first variable in this category refers to parents with the highest socio-professional level: skilled white-collar workers. The results for this variable are statistically non-significant, so the odd-ratios results do not indicate a better performance by the students whose parents belong to this group in particular. However, students whose parents belong to the non-skilled white-collar socio-professional category seem to have a consistently lower probability of obtaining low results in the four years evaluated: 20% less during 2000 and 2003, and 30% less during 2006 and 2009, compared to the base category of blue-collar workers. It is important to notice that these results are very significant for the 2006 and 2009 years.

The general tendency seems to be that when both parents work, there is a slight reduction in the possibility of obtaining a score under level-2 in reading com-

petency, the 2000 regression being an exception, as its positive coefficient suggests the opposite relation. The effect seems to be consistent through time, but it is not possible to obtain a conclusive result.

Highest educational achievement by the parents has the effect of decreasing the risk of obtaining an unsatisfactory result in reading competency. Cordero, Crespo and Pedraja (2013) conclude that the majority of the literature using PISA data for Spain found this recurrent relation between the educational attainment of the parents and the academic results of the students. The continuous variable indicates that the higher the level of education of the parents the lower the probability of being at risk. It is important to mention that although statistically significant, the odd-ratios do not indicate an important reduction in the probability of obtaining a result below level-2 in the test, probably showing that the impact of parental education operates indirectly through other variables.

Cultural and educational resources of the household

The first variable in this category (HEDR) is an index provided by PISA that measures the possession of home educational resources. Results in Table 6 show that the possession of these resources is statistically very significant and decreased the probability of obtaining a low result in reading competence by 40% in 2000, and by 30% during the 2003, 2006 and 2009 regressions.

The variable that incorporates the effect of cultural possessions within the family, CULT, gathers three elements together: whether the household owns classical literature, books of poetry or works of art[4]. The results of the regressions for this variable suggest that the presence of cultural elements in the home significantly reduces the risk of obtaining a result under level-2 in reading competence by 40% for 2000, 2003 and 2006. However in 2009 cultural possessions seemingly had less of an effect on performance in comparison with that observed for the previous years.

4 Further explanation of the construction of HEDR and CULT can be found in OECD (2000; p.225); OECD (2003; p.283); OECD (2006; p. 316) and OECD (2009 I, p.112).

School variables

School variables are divided into five categories: types of schools and their location, school inputs, school composition, environment of the schools and autonomy of the school.

School characteristics

The first two variables in this category are related to the size of the town/city where the school is located. They intend to capture whether the concentration of population has an effect on the results of the students that belong to schools in a certain area.

Schools located in municipalities with populations of over a million inhabitants present a remarkable result. The signs of the coefficients are stable and negative during the four years considered, but the effect of the variable, as the evolution of the coefficients suggests, appears to increase over time (Table 6), indicating that students from schools located in metropolitan zones have an increasingly lower probability of being at risk.

Within this category, a second group of variables considered are those that describe the schools by the type of ownership.

According to the results obtained by Calero and Waisgrais (2009), the ownership of the school appears to have a neutral effect on the probability of obtaining a deficient result in reading competency under level-2, with the sole exception of the students enrolled in private publicly-funded schools in 2006, where the risk of school failure was reduced by 30% compared to students attending state schools.

The last variable contemplated in this category, SCHLSIZE, describes the size of the school in terms of the number of students. As can be seen in Table 6, the proportion of students who perform poorly in reading competency appears to be large in schools where the number of students is under the average. However the lack of statistical significance underlines the neutral effect of the variable on the risk of school failure.

School resources

In this category two variables have been taken into consideration. The first measures the student-computer ratio. The second is the student-teacher ratio, identified by the STRATIO variable. Both variables are statistically non-significant in all the regressions. Overall, results in this category are consistent with those obtained by Calero, Choi and Waisgrais (2010).

School composition

Three variables are considered in this category. The first describes the percentage of girls in the school population. The regression results indicate a positive impact on reducing the probability of school failure risk when the percentage of girls in the student population is over the average, but the estimates are only statistically significantly in 2003.

A ratio of immigrant students over 20% of the student population (IRATIO 20) is statistically significant for both the 2000 and 2003 regressions (Table 6). The positive sign of the regression indicates that students in schools with more than 20% of immigrant students had a higher risk (130% and 150% respectively) of obtaining a result under level-2 in reading competency in 2000 and 2003.

In 2006 and 2009 the variable IRATIO20 was replaced by IRATIO30. This variable was introduced because exploratory regressions using the IRATIO20 variable showed that the results were not statistically significant for these two years. With the introduction of IRATIO30, the significance threshold of the variable was increased. This variable only considered schools in which immigrant students were over 30% of the total school population. The results show that this variable appears to be slightly significant only in 2006. These results are in line with most of the PISA-based literature for Spain, such as Cordero, Crespo and Pedraja (2013), which usually finds that the proportion of immigrant students in a school has negative effects on the academic results of the native students when it exceeds the 20% threshold.

The school environment

Schools with a large number of parents with ISCED 5 and 6 educational levels are compared to the rest of the schools. The results of the regressions indicate, as expected, that schools with a larger proportion of parents with tertiary education have a student population that appears to have a lower risk of obtaining a result below level-2 in reading competency compared to schools where the educational achievement of the parents is lower.

The other variable in this category, SPLP2, indicates schools in which a majority of families belong to the white-collar socio-professional category. The regression coefficients are negative and statistically significant in 2003 and 2006. The odd-ratios corroborate the idea that schools where there are more families whose parents are white-collar workers are characterized by students that appear to have a smaller probability of obtaining a poor result in the reading test evaluation. Students

in this group of schools were 30% less at risk in 2000, 40% in 2003 and 2006 and 20% in 2009. As can be observed, the tendency seems to be decreasing over time.

School autonomy

Two variables measure the degree of autonomy of the school in this category. The variable B_MNGMNT describes the level of school budgetary autonomy. The results of the regression are statistically significant in 2000 and 2009, at a level of 50% and 40% respectively, suggesting that a high degree of independence in the allocation of resources has an important effect in decreasing the risk of students obtaining low results in reading competency compared to schools with less autonomy.

The C_CONTENT variable denotes schools with a significant degree of autonomy in the selection of curriculum content. The lack of statistical significance implies that the variable has a neutral effect on the probability of obtaining a result under level-2 in reading competency. The regression outcomes for both school autonomy variables do not offer a conclusive result and further research is suggested for the future.

Trends over the decade

To compare the performance of students in reading competence over the 2000 decade, this final section describes the trends for a set of specific variables, tracking their evolution in PISA between 2000 and 2009. The calculation of changes in a variable between the 2000 and 2009 PISA tests will only be meaningful if its definition does not change in time. This is true for reading competency, the only competency for which the theoretical framework has remained unchanged throughout all the PISA tests (OECD, 2009a). We therefore calculate trend indicators following the three-step method suggested by the OECD (2009a) for two outcome variables: reading performance and the proportion of students below level-2.

The results in table 7 suggest that performance in reading competency by Spanish students has decreased during the 2000 decade. We have therefore split the analysis by gender, family origin and school ownership in order to provide possible explanations for this fact. Among all these subgroups, we only find a significant decrease in the performance of male students. A possible explanation for this result might be that, during the 2000 decade, non-skilled workers easily found well-paid jobs in the Spanish labor market, mainly in the real estate and service sectors. In this context, a number of young –mainly male- students might not have found incentives for continuing with their studies.

Table 7: Trends in PISA reading performance variables (2000 - 2009)

	2000		2009		2009 - 2000		
	MEAN SCORE 2000	STANDARD ERROR	MEAN SCORE 2009	STANDARD ERROR	DIFERENCE 2009 - 2000	STANDARD ERROR	STANDARDIZED DIFFERENCE
Female	505	2,8	496	2,2	-9	6,09	-1,48
Male	481	3,4	467	2,2	-14	6,38	-2,19
Total	493	2,7	481	2,0	-12	5,97	-2,01
Publics Schools	478	3,5	469	2,3	-9	6,47	-1,39
Private Schools	515	5,1	505	3,8	-10	8,05	-1,24
Native Students	494	2,6	488	2,0	-6	5,93	-1,01
1st Generation Immigrants	450	15,9	428	3,9	-22	17,10	-1,29
2nd Generation Immigrants	460	17,8	464	8,4	4	20,29	0,20

Source: OECD-PISA, 2000 and 2009 databases.

The results in table 8, which presents the trend followed by the main outcome measured in this study, the proportion of students performing below level-2 in PISA, are consistent with this view. According to these results, the proportion of students who obtained a score below level-2 between 2000 and 2009 remained unchanged, suggesting a lack of significant advances in the reduction of school failure risk throughout the decade.

Table 8: Percentage of students with scores below level-2 in reading competency

	2000		2009		2009 - 2000		
	% OF THE POPULATION	STANDARD ERROR	% OF THE POPULATION	STANDARD ERROR	DIFERENCE 2009 - 2000	STANDARD ERROR	STANDARDIZED DIFFERENCE
Scores below level-2 (female)	11,5%	1,1	14,6%	0,9	3,1%	1,42	0,02
Scores below level-2 (male)	20,4%	1,4	24,4%	1,0	4,0%	1,72	0,02
Scores below level-2 (total)	16,3%	1,1	19,6%	0,9	3,3%	1,42	0,02

Source: OECD-PISA, 2000 and 2009 databases.

Additionally, there are two PISA indexes, relevant to this study, that satisfy the necessary conditions for performing adequate time-trends estimations, as their definition remained identical between 2000 and 2009: HISEI (Highest International Socio Economic Index), and ESCS (Index of Economic, Social and Cultural Status)[5]. The results in table 9 suggest that the HISEI index, constructed around a self-reported parental occupation indicator, has remained unchanged during the 2000 decade. The ESCS index is statistically different from zero between the two samples. The trend followed by this indicator -that measures the socio-economic background based on parent's education levels and occupational status and possessions at home- seems to imply an improvement in the socio-economic conditions of households in Spain during the decade. In other words, the reading performance of Spanish students decreased during the 2000 decade although the socioeconomic situation of their households improved. However, combining this finding with those presented in tables 8 and 9, it might also be the case that the improvement in the economic situation (reflected through the ESCS status) had a negative impact by reducing the student's incentives to study. Future studies should focus on this hypothesis. Nevertheless, the reduction in school failure rates observed in Spain since 2009 –the beginning of the economic crisis- seems to reinforce the credibility of this hypothesis and, in a sense, the expression "school failure" could be seen as unfair, as school characteristics might not be the main factors explaining the decrease in reading performance observed between 2000 and 2009.

Table 9: Trends in PISA variables other than performance: HISEI and ESCS (2000 – 2009)

	2000		2009		2009 - 2000		
	HISEI 2000	STANDARD ERROR	HISEI 2009	STANDARD ERROR	DIFERENCE 2009 - 2000	STANDARD ERROR	STANDARDIZED DIFFERENCE
Highest International Socio Economic Index (HISEI)	44,99	0,6	44,40	0,5	-0,59	0,78	-0,76
Index of Economic Social and Cultural Status (ESCS) level-2 (male)	-0,56	0,05	-0,31	0,03	0,25	0,06	4,29

Source: OECD-PISA, 2000 and 2009 databases.

5 OECD (2009b) describes the items included in both indexes.

CONCLUSIONS

The most interesting contribution of this paper to the previous literature on school failure risk is the possibility of observing the evolution over a whole decade of a group of factors that appear to have an important impact on the academic performance of students in Spain.

Among the personal variables the gender of the students has a strong effect on the probability of being at risk of school failure. Girls consistently perform better in reading competency than males. The ratio 2 to 1 appears unvarying throughout the decade, indicating that male students have twice the probability of school failure risk compared to female students. Although the real importance of gender mm may be overestimated d due to the specific selection of reading competence, school failure seems to be a mainly masculine issue in Spain. As has been explained, the relationship between school failure and the labor market deserves further research.

There is also a significant difference in the results for immigrant students compared to students born in Spain. This divergence is remarkably accentuated in the risk of school failure for first generation immigrants. The accumulation of a number of pre-conditions that seem to be characteristic of these particular students suggests that immigrant students begin from a situation of enormous disadvantage compared to national students. In this sense, policies that help to ease the process of integration of immigrants into Spanish society, and policies that increase the instruction time of immigrants in the schools, could have a positive impact on the academic performance of these students.

Household characteristics offer an important insight into the influence of the family environment on academic performance. Students with highly educated parents, belonging to the white-collar socio-professional category, along with the possession of cultural and educational resources at home, show that these things have a strong positive influence on the reduction of the risk of school failure. Therefore, further efforts should be made to reduce differences in family background and educational inputs at home. Focused grants and programs could be useful tools in achieving this objective.

Students in schools where over 20% of the total school population are immigrant students and in schools with a predominance of blue-collar families face a greater risk of school failure. In schools where parents with tertiary education are predominant, the risk to their students does not seem to decrease. Policies guaranteeing a more homogeneous distribution of immigrant students among schools, keeping ratios below 20%, could have a powerful effect on the students' academic performance. The strong impact of school composition variables on academic out-

come seems to support the need to reduce the segregation of students between public and private publicly-funded schools. If the educational system is meant to provide equality of opportunity to young citizens, strategies should be studied for reducing the over-concentration of low socio-economic level, immigrant students and blue-collar families in public schools.

Among the school variables the location of the school in a large city has a positive influence on the reduction of school failure risk, perhaps because of the availability of social, cultural and educational resources in more densely populated areas. After controlling for the socio-economic characteristics of the schools, the students in private and private publicly-funded educational institutions do not appear to have a lower risk of school failure than students from state schools.

The evolution of the determinants during the ten years covered by the study shows particularly stable behavior over time for personal characteristics which become, in general, more significant during the last half of the decade. Most of the household variables are also exceptionally stable during these ten years. Parental occupation deserves a special mention, appearing positively related to the risk of school failure in 2000 and then negatively related for the rest of the years. A possible hypothesis is that there is not necessarily a reduction in the risk of school failure for the economically active households, but perhaps there is an increase in the risk for those families facing problems in the labor market due to the economic crisis.

In contrast to the first-level variables, most of the school determinants are not significant during the decade. However, this fact does not imply the nonexistence of a relation between these variables and school failure risk. One reason behind the non-significance of school resources could be the specification of the model, focused on measuring the mean effects on the population sample. A possible alternative for future research is the analysis of heterogeneous effects by considering different subgroups.

It must be acknowledged that the cross-sectional analysis of the study of school failure risk is a limitation of this research. This paper, however constrained by these restrictions, has still provided relevant information on the cumulative processes that surround school failure. Indeed it has been shown that while the proportion of students at risk of school failure has remained relatively unchanged during the 2000-2009 period, overall performance has decreased. Although unable to establish a causal relationship, our study suggests the existence of factors external to the educational system, such as the labor market situation and the sectorial structure of Spain, which may be affecting these trends.

Future studies should simultaneously include the outcomes for the three competencies and, ideally, should draw on panel data. Also, more detailed studies going

more deeply into the question of immigrant students and the difficulty of measuring peer effects would be relevant in subsequent works.

Fecha de recepción del original: 22 de enero de 2013
Fecha de aceptación de la versión definitiva: 25 de junio de 2013

REFERENCES

Aikens, N. and Barbarin, O. (2008). Socioeconomic differences in reading trajectories: The contribution of family, neighborhood, and school contexts. *Journal of Educational Psychology*, *100*(2), 235-251.

Berger, L. M., Hill J. and Waldfolgel J. (2005). Maternity Leave, Early Maternal Employment and Child Health and Development in the US. *The Economic Journal*, *115*, 29–47.

Bertrand, M. and Pan, J. (2013). The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behavior. *American Economic Journal: Applied Economics*, *American Economic Association*, *5*(1), 32-64.

Choi, Á. and Calero, J. (2013). Determinantes del riesgo de fracaso escolar en España en PISA-2009 y propuestas de reforma. *Revista de Educación*, *362*, 562-593.

Calero, J., Choi, Á. and Waisgrais, S. (2010). Determinantes del Riesgo de Fracaso Escolar en España: Una Aproximación a través de un Análisis Logístico Multinivel aplicado a PISA-2006. *Revista de Educación*, número extraordinario 2010, pp. 225-256.

Calero, J. and Escardibul, J. (2007). Evaluación de Servicios Educativos: el rendimiento en los centros públicos y privados medido en PISA-2003. *Hacienda Pública Española*, *183*, 33-66.

Calero, J. and Waisgrais, S. (2009, February). *Rendimientos Educativos de los alumnos inmigrantes: identificación de la incidencia de la condición de inmigrante y de los* peer effects. XVI Encuentro de Economía Pública. Universidad de Granada, Granada.

Carrington, B., Tymss, P. and Merrell, C. (2008). Role Models, School Improvement and the Gender Gap. Do Men Bring out the Best in Boys and Women the Best in Girls? *British Educational Research Journal*, *34*(3), 315-327.

Chingos, M. (2010). The Impact of a Universal Class-Size Reduction Policy: Evidence from Florida's Statewide Mandate. *Economics of Education Review*, *31*, 543-562.

Cohen, P. (1987, April). *A Critical Analysis and Reanalysis of the Multisection Validity meta-analysis*. Annual meeting of the American Educational Research Association 1987, Washington.

Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F. and York, R. (1966). *Equality of Education Opportunity*. Washington: U.S. Government Printing Office.

Cordero, J. M., Crespo, E. and Pedraja, F. (2013). Rendimiento educativo y determinantes según PISA: Una revisión de la literatura en España. *Revista de Educación*, *362*, 273–297.

Cordero, J. M., Manchón, C. and García, M. Á. (2011). Los resultados educativos españoles en PISA 2009 y sus condicionantes. *Investigaciones de Economía de la Educación*, *6*, 70-87.

Crawford, C., Dearden, L. and Greaves, E. (2011). *Does when you are born matter? The impact of month of birth on children's cognitive and non-cognitive skills in England*. Institute for Fiscal Studies. Retrieved November 8, 2012 from http://www.ifs.org.uk/bns/bn122.pdf

Entorf, H. and Minoiu, N. (2005). What a Difference Immigration Policy Makes: A Comparison of PISA Scores in Europe and Traditional Countries of Immigration. *German Economic Review*, *6* (3), 355-376.

Ferguson, R. and Ladd, H. (1996). How and Why Money Matters: An Anlysis of Alabama Schools. In R. Ferguson and H. Ladd, *Holding Schools Accountable. Performance-Based Reform in Education* (pp. 265-298). Washington D.C.: Brookings Institution Press.

Fernández, E., Mena, L. and Riviere, J. (2010). *School Failure and Dropouts in Spain.* Retrieved October 3, 2011 from http://obrasocial.lacaixa.es/deployedfiles/obrasocial/Estaticos/pdf/Estudios_sociales/vol29_en.pdf

García, J. (2011). Nivel Socioeconómico, tipo de Escuela y Resultados Educativos en España: El Caso de TIMSS PIRLS 2011. (Vol. II, pp. 66-104) Retrieved January 21, 2012 from http://www.mecd.gob.es/dctm/inee/documentos-de-trabajo/dtmontalvo.pdf?documentId=0901e72b814db8c1

Hanushek, E. A. (1997). Assessing the effects of School Resources on Student Performance: An Update. *Educational Evaluation and Policy Analysis*. *19*, 141-64.

Hanushek, E. A. (2003). *The* Failure of Input-Based Schooling Policies. *The Economic Journal*, *13*, 64-98.

Haveman, R. and Wolfe B. (1995). The Determinants of Children's Attainments: A Review of Methods and Findings. *Journal of Economic Literature*, *American Economic Association*, *33*(4), pp. 1829-1878.

Holmes, C. T. (1989). Grade level retention effects: A meta-analysis of research studies. In L. A. Shepard and M. L. Smith (Eds.), *Flunking grades: Research and policies on retention*, (pp. 16-33). Philadelphia: The Falmer Press.

Instituto Vasco de Investigación y Evaluación Educativa (2009). *Evaluación Diag-*

nostica 2009. Análisis de Factores y Variables. 2º Educación Secundaria Obligatoria. Retrieved May 15, 2011 from http://isei-ivei.net/cast/pub/ED09_inf_variables/ED09_2ESO_variables.pdf

Jimerson, S. R. (2001). Meta-analysis of Grade Retention Research: Implications for Practice in the 21st Century. *School Psychology Review*, *30*(3), 420-437.

Krueger, A. B. (2002). Economic Considerations and Class Size. *The Economic Journal*, *113*(485), 34-63.

Lavy, V. (2012). *Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing countries*. Retrieved August 17, 2011 from http://www.nberg.org/papers/w16227

Martin, A. (2009). Age Appropriateness and Motivation, Engagement, and Performance in High School: Effects of Age within Cohort, Grade Retention, and Delayed School Entry. *Journal of Educational Psychology*, *101*(1), 101-114.

McCarthy, K. (1998). *Adaptation of Immigrant Children to the United States: A Review of the literature*. New Jersey: Center for Research on Child Wellbeing - Princeton University.

Ministerio de Educación (2009). *Panorama de la educación. Indicadores de la OCDE 2009*. Retrieved May 18, 2010 from http://www.mecd.gob.es/dctm/ievaluacion/internacional/indicadores-panorama-de-la-educacion-2009.pdf?documentId=0901e72b8046dc8d

Organisation for Economic Co-operation and Development (2000). *PISA 2000: Technical Report PISA 2000*. Retrieved August 10, 2011 from http://www.oecd.org/pisa/pisaproducts/33688233.pdf

Organisation for Economic Co-operation and Development (2003). *PISA 2003: Technical Report PISA 2003*. Retrieved August 10, 2011 from http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/35188570.pdf

Organisation for Economic Co-operation and Development (2006). *PISA 2006: Technical Report PISA 2006. Retrieved* August 10, 2011 from http://www.oecd.org/pisa/pisaproducts/42025182.pdf

Organisation for Economic Co-operation and Development (2009a*). PISA Data Analysis Manual. SPSS Second Edition*. Retrieved August 10, 2011 from http://browse.oecdbookshop.org/oecd/pdfs/free/9809031e.pdf

Organisation for Economic Co-operation and Development (2010). *Education at Glance 2010: OECD Indicators*. Retrieved April 22, 2011 from http://www.oecd.org/education/skills-beyond-school/45926093.pdf

Organisation for Economic Co-operation and Development (2011). *Education at*

Glance 2011: OECD Indicators. Retrieved July 7, 2012 from http://www.oecd.org/education/skills-beyond-school/48631582.pdf

Organisation for Economic Co-operation and Development (2010). *PISA 2009 Results: Learning to Learn: Student Engagement, Strategies and Practices*. Retrieved July 17, 2012 from http://www.oecd.org/pisa/pisaproducts/48852630.pdf

Pfeffermann, D., Skinner, C. J., Holmes D., Goldstein, H. and Rasbash, J. (1998). Weighting for Unequal Selection Probabilities in Multilevel Models. *Journal of the Royal Statistic Society*, B, *60*, 123-40.

Raudenbusch, S. and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Bethesda, Maryland: Sage Publications Inc.

Reynolds, A. and Temple, J. (1998). Extended Early Childhood Intervention and School Achievement: Age 13 findings from the Chicago Longitudinal Study. *Child Development*, 69(1), 231-246.

Robertson, E. (2011). The Effects of Quarter of Birth on Academic Outcomes at the Elementary School Level. *Economics of Education Review*, *30*, 300-311.

Rubin, D.B. (1987). *Multiple Imputation for non-response in surveys*. New York: John Wiley & Sons.

Schleicher, A. (2007). Can competencies assessed by PISA be considered the fundamental school knowledge 15-year-olds should possess? *Journal of Educational Change*, *8*(4), 349-357.

ANNEX A: VARIANCE REDUCTION ANALYSIS RESULTS

2000	VARIANCE COMPONENT	STANDARD DEVIATION	CHI-SQUARE	P-VALUE
Null model	0,845	0,919	871,38	0,000
Full model	0,326	0,569	413,72	0,000
2003				
Null model	0,855	0,925	1479,38	0,000
Full model	0,478	0,690	1005,15	0,000
2006				
Null model	0,838	0,916	2716,51	0,000
Full model	0,465	0,681	2055,64	0,000
2009				
Null model	1,000	1,000	4697,29	0,000
Full model	0,583	0,763	3246,37	0,000

Note: The table reports the difference between the variance component for the unconditional model with random intercept (one-way ANOVA) and the full model. The table reports information about the outcome variability of within-group and between-group variance. The significant difference in the variance between groups in the four estimations justifies the use of hierarchical models.

La evaluación del aprendizaje de estudiantes: validación española del Assessment Experience Questionnaire (AEQ)

The Assessment of Student Learning: Spanish Validation of the Assessment Experience Questionnaire (AEQ)

JUAN L. NÚÑEZ
Universidad de Las Palmas de Gran Canaria
juanluis.nunez@ulpgc.es

CARMEN I. REYES
Universidad de Las Palmas de Gran Canaria
creyes@dedu.ulpgc.es

Resumen: El propósito de esta investigación fue traducir al español y analizar las propiedades psicométricas del Assessment Experiences Questionnaire (AEQ) ya que no se dispone de un cuestionario en español que evalúe la experiencia de evaluación del aprendizaje de los estudiantes. Los participantes fueron 329 estudiantes universitarios. Los resultados del análisis de correlaciones y de la regresión lineal múltiple apoyaron la validez predictiva. El análisis factorial confirmatorio confirmó la estructura de nueve factores. Además, la consistencia interna fue aceptable y la correlación test-retest fue moderada. En conclusión, los resultados apoyan preliminarmente el uso de la versión española del AEQ.

Palabras clave: evaluación orientada al aprendizaje; propiedades psicométricas; educación superior; rendimiento académico.

Abstract: The aim of this study was the translation of the Assessment Experiences Questionnaire (AEQ) into Spanish, and the analysis of its psychometric properties. We carried out this study because there is no questionnaire in Spanish which evaluates the experience of learning assessment. Participants were 329 university students. Results of the correlation analysis and the multiple linear regression supported the predictive validity of the questionnaire. The confirmatory factor analysis confirmed the nine-factor structure. In addition, internal consistency was acceptable, and the test-retest correlation was moderate. In conclusion, the results support the use of the Spanish version of the AEQ as an instrument to assess the students' perception of the experience of learning assessment.

Keywords: learning-oriented assessment; psychometric properties; higher education; academic performance.

INTRODUCCIÓN

A partir de los años 90 en determinados ámbitos investigadores se inicia un cambio en la concepción sobre la naturaleza del aprendizaje cuyo foco de atención se centrará en la capacidad de un estudiante de articular y utilizar globalmente los conocimientos, destrezas y actitudes necesarios para solucionar problemas específicos (formación basada en competencias) dejando atrás el modelo tradicional preocupado por lo que sabe un alumno en cuanto contenido.

Esto originará un cambio en el marco conceptual de la evaluación educativa desplazando el interés casi exclusivo de principios psicométricos hacia otros centrados en la evaluación educativa (Mateo, 2006). Este nuevo enfoque alternativo de la evaluación es muy amplio y en él podemos encontrar diferentes modalidades o perspectivas: evaluación auténtica, evaluación orientada al aprendizaje, evaluación del desempeño o ejecución, etc. Es viable descubrir algunas diferencias entre el significado de estos términos pero todas estas perspectivas tienen el siguiente objetivo fundamental: enfatizar que el propósito fundamental de la evaluación debería ser promocionar el aprendizaje.

En adelante nos referimos a este enfoque como evaluación orientada al aprendizaje (EOA) porque estos vocablos destacan una de las características principales: la importancia de la evaluación en la optimización del aprendizaje. Este término es utilizado por primera vez en educación superior por Carless (2003) quien en diferentes trabajos ha señalado las principales condiciones de esta conceptualización de la evaluación en la universidad (Carless, 2007, 2009; Carless, Salter, Yang y Lam, 2011): a) las tareas de evaluación son entendidas como tareas de aprendizaje. Esto quiere decir que las tareas de evaluación deben promover experiencias de aprendizaje profundo que se orienten hacia los resultados de aprendizaje esperados; b) la implicación de los estudiantes en la evaluación de tal forma que desarrollen una mejor comprensión de las metas de aprendizaje y que se ocupen más activamente en los estándares y criterios y c) los estudiantes deben recibir retroalimentación o feedback adecuado que podrían utilizar como prealimentación o feedforward en el trabajo futuro.

Desde la EOA, Biggs (2005) enfatiza la importancia del alineamiento constructivo del currículo es decir, la evaluación debe estar alineada con los resultados de aprendizaje y las actividades a realizar, si realmente queremos desarrollar una educación de calidad. Recientemente McDowell, Wakelin, Montgomery y King (2011) aportan una variación a la conceptualización de la EOA a partir de un modelo comprensivo que no sólo se centra en la calidad y oportunidad de feedback que le aportan los tutores o del diseño de tareas (aunque ambos sean importantes),

sino en la necesidad que tienen los estudiantes de un ambiente que les permita valorar cómo lo están realizando y ofrecerles oportunidades para mejorarlos. Han definido la EOA como un ambiente de evaluación que es rico en feedback formal, por ejemplo comentarios del tutor, sistemas de autoevaluación, etc.; es rico en feedback informal a través del dialogo equitativo y equilibrado en el aula y la interacción de pares provee oportunidades para probar y practicar conocimientos, comprensión y destrezas; desarrolla tareas de evaluación que son auténticas o relevantes de cara a la actividad profesional; ayuda a los estudiantes a desarrollar autonomía e independencia y presenta un equilibrio apropiado entre evaluación formativa y sumativa.

Desde este enfoque, se asume que la evaluación afecta a la forma de estudiar de los aprendizajes (Gibbs, 2006). La evaluación enmarca el aprendizaje, genera la actividad de aprendizaje y orienta la mayor parte de los aspectos de la conducta de aprendizaje (Gibbs y Simpson, 2004). Todas las percepciones de la evaluación de los estudiantes influyen considerablemente en las aproximaciones al aprendizaje y a la realización de los estudiantes. Turner y Gibbs (2010) han confirmado que las respuestas de aprendizaje de los estudiantes en ambientes diferentes de evaluación pueden ser notablemente diferentes. Las respuestas positivas de aprendizaje (aproximación al aprendizaje profundo, cobertura total del programa, aprendizaje desde la preparación de un examen) se ha mostrado que están asociadas con mucha evaluación formativa, poca evaluación sumativa y mucho feedback oral. Se ha encontrado que las respuestas negativas de aprendizaje (ser estratégico en cuanto estudiar el programa y en una aproximación superficial) está asociado a mucha y variada evaluación sumativa, detallada especificación de resultados y criterios y poca evaluación formativa. Por lo que podemos decir que hay evidencias que confirman que existe una relación entre el tipo de ambiente de evaluación adoptado por una universidad y la actividad de estudio de sus estudiantes.

En síntesis, podríamos concebir la EOA como un constructo teórico que se centra en la evaluación, como un proceso interrelacionado con el aprendizaje basado en la implicación de los estudiantes en tareas apropiadas, en la provisión de abundante feedback incluyendo recursos de auto y coevaluación y sustentado en un compromiso por mejorar el aprendizaje por todos. Esta modificación en la conceptualización de la evaluación en las investigaciones también está generando numerosas innovaciones educativas, aunque en la práctica cotidiana de las universidades españolas la evaluación sigue teniendo un carácter tradicional centrada en la calificación, dirigida por el profesor y donde la participación de los estudiantes es prácticamente nula (Ibarra y Rodríguez, 2010; Reyes, Sosa, Marchena y Marchena, 2012).

En nuestro país, observamos que a pesar que, de alguna manera, se están planteando algunas mejoras, necesitamos instrumentos formales que nos proporcionen una visión y un panorama más amplio sobre las formas en las que las experiencias de evaluación de los estudiantes están influyendo en el aprendizaje. Desde este marco conceptual, las tareas o experiencias de evaluación constituyen el elemento que más influye sobre el aprendizaje. De ahí, la necesidad de analizar cómo están diseñadas: tiempo de esfuerzo del alumno, temas en que se centran, etc. y, de esta manera, detectar, si realmente están apoyando el aprendizaje.

Actualmente, prácticamente todas las universidades están sumergidas en procesos de calidad en donde, a través de cuestionarios, entre otros instrumentos, se analiza la calidad de la enseñanza universitaria: claridad de los objetivos, carga de trabajo apropiado, evaluación adecuada, etc. Pese a esto, no disponemos de un cuestionario que, desde esta perspectiva, nos ayude a diagnosticar el impacto de las experiencias de evaluación a fin de que los profesores puedan utilizarlo para evaluar el diseño de sus sistemas de evaluación con objeto de realizar mejoras orientadas a la optimización del aprendizaje.

En el contexto anglosajón podemos encontrar una larga lista de inventarios relacionados con el aprendizaje de los estudiantes con sus experiencias en el curso y módulos (Richarson, 2005). Entre todos ellos, el más reciente y ampliamente conocido en Educación Superior en el Reino Unido es el Assessment Experience Questionnaire (AEQ) desarrollado por Gibbs y Simpson (2003) y que ha sido traducido a diversos idiomas. Este cuestionario supone una herramienta para que el profesor pueda diagnosticar hasta qué punto la evaluación de su asignatura apoya el aprendizaje de sus estudiantes y así, poder establecer cambios en la evaluación. Los items y las subescalas del AEQ se desarrollaron a partir de una revisión teórica de la literatura relativa a la promoción del aprendizaje por medio de la evaluación. Esta revisión permitió identificar una serie de condiciones bajo las cuales la evaluación fomentaba el aprendizaje, que se agruparon en cinco apartados: cantidad y distribución del esfuerzo, calidad y nivel de esfuerzo, cantidad y duración del feedback, calidad del feedback y respuesta del estudiante al feedback. Posteriormente, se realizaron entrevistas abiertas a estudiantes sobre su experiencia de evaluación, que dieron lugar a un conjunto potencial de items relacionados con los cinco apartados citados anteriormente. A partir de aquí, los autores desarrollaron un cuestionario piloto de seis subescalas de seis items cada una. Este cuestionario se administró a dos muestras de dos instituciones diferentes, 1050 y 529 estudiantes, respectivamente. Los resultados de un análisis factorial exploratorio mostraron una solución de seis factores que explicaban el 50% de la varianza y que no coincidían plenamente con los apartados resultantes de la revisión teórica: demanda del tiempo y

distribución del esfuerzo, tareas y aprendizaje, cantidad y duración del feedback, calidad del feedback, qué hacer con el feedback y aprendizaje desde el examen. La consistencia interna medida con el alfa de Cronbach osciló entre .44 y .87.

Con el objetivo de mejorar las propiedades psicométricas del AEQ, Gibbs y Dunbar-Goddet (2007) desarrollaron una versión revisada del cuestionario denominada AEQ (V3.3) que contenía cinco escalas del AEQ (cantidad de esfuerzo, cobertura del programa, cantidad y calidad de feedback, uso del feedback y aprendizaje desde el examen), dos escalas del Course Experience Questionnaire (CEQ; Ramsden, 1991), dos escalas adicionales (aproximación profunda y aproximación superficial al estudio) y un item de satisfacción general. Esta versión de 28 items fue administrada a 516 estudiantes. El análisis factorial exploratorio mostró una estructura coherente con pesos factoriales por encima de .50 y valores alfa de las subescalas entre .61 y .85. Además, el cuestionario fue capaz de discriminar entre dos planes de estudio diferentes.

Puesto que la percepción que tienen los estudiantes de la experiencia de evaluación del aprendizaje es importante para la mejora de los procesos de calidad en educación superior y no se dispone de un cuestionario en español que evalúe esta experiencia, nuestro objetivo es traducir al español y analizar las propiedades psicométricas del cuestionario AEQ (V3.3). Para ello, se llevará a cabo un análisis descriptivo de los ítems del AEQ (V3.3), un análisis de correlaciones y un análisis de regresión lineal múltiple por pasos entre los factores que configuran la escala y el rendimiento académico como variable criterio para analizar la validez predictiva; además, se analizará la evidencia de validez de constructo a través de un análisis factorial confirmatorio y la evidencia de fiabilidad a través de la consistencia interna y la estabilidad temporal.

MATERIAL Y MÉTODOS

Participantes

Los participantes fueron 329 estudiantes universitarios de distintas titulaciones de la Universidad de Las Palmas de Gran Canaria, 286 mujeres y 43 hombres, con una media de edad de 21. 70 años (*DT* = 3.80). El tipo de muestreo fue por conglomerado donde la unidad de análisis fue el aula.

Para el análisis de la estabilidad temporal del cuestionario se utilizó una segunda muestra de 56 estudiantes universitarios, 21 hombres y 35 mujeres, con una media de edad de 19.91 años (*DT* = 2.36) que cumplimentaron el cuestionario en dos ocasiones tras un intervalo de cuatro semanas.

Instrumentos

Para medir la experiencia de evaluación del aprendizaje de los estudiantes se utilizó la versión 3.3 del Assessment Experience Questionnaire (AEQ; Gibbs y Dunbar-Goddet, 2007).

Para traducir este cuestionario al español se adoptó la estrategia de traducción inversa (Hambleton, 1996). Se tradujeron los ítems al español y posteriormente otro grupo de traductores volvió a traducirlos al inglés, comparándolos con los originales. Posteriormente, se sometieron los ítems a una evaluación por parte de tres expertos en la temática (Lynn, 1986) que estimaron que los ítems fueron pertinentes para medir el constructo de interés, además de la correcta redacción de los mismos en cuanto a la utilización de la gramática española. Este cuestionario está compuesto por 27 ítems (e.g. utilicé el feedback que recibí para revisar lo que había hecho en mi trabajo) distribuidos en nueve subescalas que miden la experiencia de evaluación del aprendizaje de los estudiantes: cantidad de esfuerzo (ítems 6 y 13), cobertura del programa (ítems 4, 5, 11 y 17), cantidad y calidad de feedback (ítems 3, 15 y 16), uso del feedback (ítems 1, 2 y 8), evaluación apropiada (ítems 10, 14 y 18), claridad de los objetivos y criterios (ítems 7, 9 y 12), aproximación profunda (ítems 20, 21 y 22), aproximación superficial (ítems 19, 23 y 24) y aprendizaje desde el examen (ítems 25, 26 y 27). Además, se incluye un item que evalúa la satisfacción en general del estudiante con el curso. Se mantuvo la distribución original de los ítems del cuestionario y la redacción en sentido negativo de 9 items (ítems 3, 5, 10, 11, 12, 14, 15, 16 y 18). Las respuestas están expresadas en una escala tipo Likert de 1 (*totalmente en desacuerdo)* a 5 (*totalmente de acuerdo)*.

Además, se solicitó a los participantes la nota media final obtenida en las asignaturas presentadas el curso anterior 2010/2011 como medida de rendimiento académico.

Procedimiento

Se requirió la autorización de los profesores, se contactó con los estudiantes durante la jornada académica y se les explicó los objetivos de la investigación. El instrumento se administró en las aulas. Se comunicó a los estudiantes que la participación era voluntaria y confidencial, enfatizando que no había respuestas correctas o incorrectas e instándoles a que la contestaran con honestidad basándose en su experiencia de evaluación del curso anterior 2010/2011. El tiempo empleado para responder fue de aproximadamente de unos 10 minutos.

Análisis de datos

En primer lugar, se llevó a cabo un análisis descriptivo de los ítems del AEQ para determinar el cumplimiento de normalidad univariada en el que se utilizó como criterio que la asimetría se situara por debajo del valor 2 y la curtosis por debajo del valor 7 (Curran, West y Finch, 1996). En segundo lugar, se realizaron un análisis de correlaciones de los nueve factores que configuran el cuestionario y la nota media final obtenida por los participantes el curso anterior 2010/2011 y un análisis de regresión lineal múltiple como evidencias de validez predictiva. En tercer lugar, se llevó a cabo un análisis factorial confirmatorio para determinar la validez de constructo del AEQ. Finalmente, se analizó la evidencia de fiabilidad a través de la consistencia interna y de la estabilidad temporal. Para realizar estos análisis se utilizaron los programas estadísticos LISREL 8.54 y PASW 18.

RESULTADOS

Análisis descriptivos

La Tabla 1 presenta los estadísticos descriptivos (media, desviación típica, asimetría y curtosis) de los ítems que configuran el cuestionario AEQ. Como se observa, todos los valores de asimetría y curtosis cumplen con el criterio de normalidad propuesto por Curran et al. (1996), lo que indica semejanza con la curva normal.

Tabla 1: Estadísticos descriptivos de los ítems de la versión española del cuestionario AEQ

ITEM	M	DT	ASIMETRÍA	CURTOSIS
1. Utilicé el feedback que recibí para revisar lo que había hecho en mi trabajo.	3.24	.66	.14	1.52
2. El feedback que recibí me hizo volver a consultar el material visto en el curso.	3.13	.67	.16	1.08
3. Apenas recibí feedback sobre mi trabajo.	3.67	.84	-.63	.61
4. Tuve que estudiar todos los temas del programa para realizar bien la evaluación.	3.55	.93	-.113	-.64
5. El sistema de evaluación me permitió ser bastante selectivo acerca de qué temas del programa debía estudiar.	3.33	.85	.03	-.42
6. Por la forma en la que se planteó la evaluación, tenías que planificarte unas horas de estudio semanales.	3.70	.83	-.02	-.51

ITEM	M	DT	ASIMETRÍA	CURTOSIS
7. Fue fácil saber los criterios de evaluación esperados.	3.18	.84	-.13	.36
8. Presté atención al feedback que me dieron los profesores sobre mi trabajo y traté de comprenderlo.	3.52	.83	.08	- .05
9. Los profesores aclararon desde el principio lo que esperaban de sus estudiantes.	3.19	.86	-.16	.26
10. El profesorado parecía más interesado en valorar lo que había memorizado que lo que había comprendido.	3.36	1.04	-.66	-.10
11. Pude ser bastante estratégico y dejar de estudiar algunos temas.	3.63	.85	-.29	-.35
12. A veces fue difícil saber lo que se esperaba de mí en este curso.	3.25	.86	-.21	-.07
13. En este curso fue necesario trabajar de forma constante para cumplir los requisitos de la evaluación.	3.85	.87	-.12	-.96
14. El profesorado muchas veces me preguntó sólo cuestiones de datos memorísticos.	3.23	.94	-.59	-.05
15. No comprendí algunos aspectos del feedback que me dieron los profesores sobre mi trabajo.	3.41	.77	-.41	.49
16. Todo el feedback que me dieron sobre mi trabajo llegó demasiado tarde para ser útil.	3.67	.79	-1.08	1.92
17. Por la forma de evaluar las asignaturas de este curso tenía que estudiar todos y cada uno de los temas.	3.26	.96	.16	-.47
18. Para tener éxito en este curso todo lo que se necesitaba era tener memoria.	3.44	.99	-.66	.029
19. Cuando leía trataba de memorizar los datos importantes que podrían ser útiles más adelante.	3.30	.93	-.13	-.173
20. Generalmente me proponía comprender a conciencia el significado de lo que me pedían que leyera.	3.62	.78	.10	-.32
21. Normalmente me esforzaba en comprender lo que inicialmente parecía difícil.	3.70	.77	-.05	-.22
22. Durante el curso, a veces me cuestioné cosas que se decían en clase o que leía.	3.47	.83	.08	-.22
23. Tenía que concentrarme en memorizar gran cantidad de información que tenía que aprender.	3.14	.94	.22	-.44

ITEM	M	DT	ASIMETRÍA	CURTOSIS
24. A veces tenía que estudiar cosas sin tener oportunidad de comprenderlas.	3.13	.97	.27	-.41
25. Hacer los exámenes me ayudó a relacionar e integrar los temas.	3.04	.88	.07	-.01
26. Aprendí cosas nuevas mientras preparaba los exámenes.	3.46	.89	-.05	-.29
27. Comprendí mejor los temas después de haber realizado los exámenes.	3.07	.95	.18	-.41
28. En general, estoy satisfecho con la enseñanza de este curso.	3.52	.86	-.17	.23

Análisis de correlaciones

Se analizó la relación entre las nueve subescalas del AEQ y el rendimiento académico como variable criterio utilizando el coeficiente de correlación de Pearson. Tal y como se observa en la Tabla II, los coeficientes son bajos, si bien, la nota media final obtenida por los estudiantes en el curso se relaciona de forma positiva y significativa con el uso del feedback recibido (r = .21; p < .01), la cobertura del programa (r = .16; p < .01), la claridad de los objetivos (r = .15; p < .01) y la aproximación profunda (r = .14; p < .05).

Tabla 2: Correlaciones entre las subescalas del AEQ y la nota media y alfas en la diagonal

	1	2	3	4	5	6	7	8	9	10
1. Nota media		.10	.16**	.06	.21**	.07	.15**	-.08	.14*	.03
2. Esfuerzo		.73	.21**	-.05	.26**	-.18**	.05	.25**	.31**	.13*
3. Programa			.70	-.15**	.03	-.15**	-.07	.13*	.15**	.05
4. Cant. feb				.74	.19**	.43**	.17*	-.39**	-.18**	.13*
5.Uso feb					.72	.15**	.28**	-.02	.21**	.30**
6. Evaluación						.73	.24**	-.58**	-.20**	.14*
7. Objetivos							.71	-.11*	.05	.11
8. Aprox.Superf.								.70	.26**	.09
9. Aprox. Prof.									.72	.18**
10. Ap. Examen										.74*

p < .05; **p < .01.

Análisis de regresión lineal múltiple

El análisis de correlaciones permite una primera aproximación a la validez predictiva que se estudia con más precisión mediante el análisis de regresión múltiple. Llevamos a cabo un análisis de regresión lineal múltiple por pasos entre las nueve subescalas del AEQ (variables independientes) y el rendimiento académico (variable dependiente), tomando la nota media de las calificaciones obtenidas por los estudiantes en las asignaturas presentadas en el curso académico 2010/2011.

Los resultados obtenidos a partir del análisis de regresión realizado muestran que, de las nueve subescalas, tres entran en la ecuación y, por tanto, tienen la capacidad de predecir de forma positiva y significativa ($p < .05$) el rendimiento académico de los estudiantes, en concreto: uso del feedback ($\beta = .18$), cobertura del programa ($\beta = .17$) y claridad de objetivos ($\beta = .12$). Los resultados obtenidos en el análisis de varianza de la regresión muestran un valor de F significativo ($F = 9.50$; $p < .05$). El coeficiente R^2 fue de .12, lo que supone que el conjunto de estas tres subescalas explica el 12% de la varianza de las calificaciones académicas.

Análisis factorial confirmatorio

La estructura factorial de la escala AEQ fue evaluada con un análisis factorial confirmatorio utilizando el procedimiento estándar de mínimos cuadrados ponderados diagonalizados y la matriz de correlaciones policóricas como entrada para el análisis de datos puesto que los ítems son variables categóricas o discretas que se responden en una escala tipo Likert (Flora y Curran, 2004). Se utilizó este método de estimación ya que no tiene limitaciones respecto al tamaño de la muestra y no requiere normalidad multivariada (Olatunji et al., 2007). Los índices de ajuste obtenidos fueron los siguientes: S – B $\chi^2/gl = 1.93$ (S – B χ^2 (288, N = 329) = 555.35, $p = .00$); IFI = .90; SRMR = .06 y RMSEA = .05 (LO 90 = .04; HI 90 = .06). Con respecto a los pesos factoriales de los ítems, oscilaron entre .43 del ítem 5 "el sistema de evaluación me permitió ser bastante selectivo acerca de qué temas del programa debía estudiar" y .86 del ítem 21 "normalmente me esforzaba en comprender lo que inicialmente parecía difícil", siendo todos ellos significativos ($p < .05$).

Fiabilidad

La consistencia interna de las nueve subescalas de la AEQ fue evaluada a través del alfa de Cronbach. Así, podemos observar en la diagonal de la Tabla 2 que los valores obtenidos se situaron entre el .70 de las subescalas cobertura del programa y

aproximación superficial y el .74 de las subescalas cantidad y calidad de feedback y aprendizaje desde el examen.

Para evaluar la estabilidad temporal de la AEQ se utilizó una muestra de 56 estudiantes de educación superior que cumplimentaron la escala en dos ocasiones con un intervalo temporal de cuatro semanas. Se determinó la estabilidad temporal mediante una prueba no paramétrica como el coeficiente de correlación de Spearman, ya que el tamaño de la muestra no permite utilizar una prueba más robusta. Los coeficientes de correlación de Spearman entre las puntuaciones del test y del retest oscilaron entre .52 de la subescala evaluación apropiada y .79 de la subescala aproximación profunda.

DISCUSIÓN

La EOA destaca la importancia de la evaluación en la optimización del aprendizaje. El cuestionario AEQ fue desarrollado para examinar el impacto de la evaluación sobre la experiencia de aprendizaje de los estudiantes. Ante la ausencia de una medida de estas características validada al español, el objetivo del presente estudio fue traducir al español y analizar las propiedades psicométricas del cuestionario AEQ en una muestra de estudiantes universitarios.

La traducción se realizó siguiendo la estrategia de traducción inversa en la misma línea que estudios anteriores (Martín-Albo, Núñez y Navarro, 2009; Núñez, Martín-Albo y Navarro, 2005).

El análisis de correlaciones entre las nueve subescalas del AEQ y el rendimiento académico como variable criterio mostró relaciones positivas entre el rendimiento y todas las subescalas, excepto con la subescala aproximación superficial. En este sentido, unos altos niveles de aprendizaje basados en la memorización de contenidos y no tanto en la comprensión de los mismos se relacionan con una nota académica baja. El análisis de regresión lineal múltiple por pasos mostró que las subescalas uso del feedback, cobertura del programa y claridad de objetivos tienen capacidad para predecir el rendimiento académico explicando el 12% de las calificaciones académicas. Estos resultados apoyan parcialmente la validez predictiva del cuestionario AEQ y revelan que deben considerarse otras variables académicas de interés, además de la experiencia de evaluación del aprendizaje de los estudiantes, como por ejemplo el clima que genera un profesor en el aula, el tipo de motivación del estudiante o el uso de estrategias cognitivas y metacognitivas para explicar en mayor medida una variable como el rendimiento académico; asimismo, ponen de manifiesto la importancia que tiene para el estudiante tener claros desde el principio los objetivos del programa, la cantidad de temas que se estudian y las respues-

tas positivas de aprendizaje (Turner y Gibbs, 2010), estando en consonancia con diversas investigaciones que destacan la influencia del feedback sobre el rendimiento de los estudiantes (Hattie, 1987, Hattie y Timperley, 2007). En términos parecidos, Black y Wiliam (1998) subrayan los efectos positivos del feedback sobre el aprendizaje y el rendimiento en los diferentes niveles de educación.

Para determinar la validez de constructo se analizó la estructura factorial del cuestionario AEQ mediante un análisis factorial confirmatorio. Los resultados confirmaron la estructura de nueve factores correlacionados, lo que corrobora la estructura propuesta por Gibbs y Dunbar-Goddet (2007) a través de un análisis factorial exploratorio. Además, los pesos factoriales de los ítems se situaron por encima de .43.

En general, los resultados obtenidos mostraron que las subescalas del AEQ presentan unos valores aceptables de consistencia interna de acuerdo a los criterios de George y Mallery (2003) y, en general, más altos que los obtenidos por Gibbs y Simpson (2003) en la primera versión del cuestionario y por Gibbs y Dunbar-Goddet (2007) en la versión 3.3. Los valores obtenidos en la correlación test-retest del cuestionario muestran una estabilidad temporal que podemos calificar de moderada-alta.

Debemos tener en cuenta algunas limitaciones. Los participantes de este estudio fueron estudiantes universitarios, por lo que se deberían analizar las propiedades psicométricas del cuestionario AEQ en otros niveles académicos. Asimismo, las evidencias de validez y fiabilidad deberían ser consideradas como provisionales dado que el tamaño de la muestra, sobre todo en el caso de los hombres, fue pequeño. Futuros estudios deberían comprobar los efectos de género, así como analizar la invarianza del cuestionario en función del género y el poder discriminante del instrumento en diferentes entornos de evaluación (e.g. evaluación formativa, evaluación sumativa, etc.). En próximas investigaciones se debería considerar el uso de ítems redactados en negativo; la redacción en positivo favorece la comprensión del ítem y podría mejorar el comportamiento psicométrico de algunas subescalas como cantidad y calidad de feedback y evaluación apropiada. Además, sería interesante desarrollar una versión reducida del cuestionario de acuerdo a las recomendaciones de Marsh, Martin y Jackson (2010) para seleccionar ítems (peso factorial, correlación ítem-test, alfa de Cronbach e importancia teórica del ítem) con la finalidad de reducir el tiempo de aplicación con las mismas garantías psicométricas.

En conclusión, los resultados mostraron que la versión española del cuestionario AEQ presenta evidencias de validez y fiabilidad, en la misma línea que los resultados obtenidos en las versiones de Gibbs y Simpson (2003) y de Gibbs y Dunbar-Goddet (2007). Por tanto, la versión española puede ser considerada como una adaptación preliminar y los resultados justifican su utilización en el contexto de

la educación superior con distintos fines. En primer lugar, podría ser utilizado como herramienta de evaluación para examinar las experiencias de evaluación de los estudiantes e indicar si éstas están orientadas al aprendizaje. Posteriormente, esto podría conducir a mejoras y cambios en el diseño de las asignaturas. Además, también podría poner de manifiesto las diferencias entre los diferentes contextos de asignaturas. Asimismo, el cuestionario AEQ podría ser utilizado para evaluar cuestiones específicas como, por ejemplo, el feedback o la claridad de los objetivos e identificar áreas donde existan lagunas y, en consecuencia, donde deberían introducirse mejoras. Por último, también podría servir como instrumento de investigación para explorar en mayor profundidad las relaciones entre el aprendizaje de los estudiantes y las características de sus ambientes o climas de aprendizaje o las relaciones entre las experiencias de EOA y las características de los estudiantes.

Fecha de recepción del original: 28 de enero de 2013
Fecha de aceptación de la versión definitiva: 14 de abril de 2013

Referencias

Biggs, J. (2005). *Calidad del aprendizaje universitario*. Narcea: Madrid.

Black, P. y William, D. (1998). Assessment and classroom learning. *Assessment in Education*, *5*(1), 7- 74.

Carless, D. (2003, Noviembre). *Learning-oriented assessment*. Comunicación presentada en la Evaluation and Assessment Conference, University of South Australia, Adelaide.

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, *44*(1), 57-66.

Carless, D. (2009). Learning-oriented assessment: Principles, practice and a project. En L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston y M. Rees (Eds.), *Tertiary Assessment & Higher Education Student Outcomes: Police, Practice & Research Learning-oriented assessment* (pp.79-90). Wellington, New Zealand: Ako Aotearoa.

Carless, D., Salter, D., Yang, M. y Lam J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*, *36*(4), 395-407.

Curran, P. J., West, S. G. y Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*(1), 16-29.

Flora, D. B. y Curran, P. J. (2004). An empirical evaluation of alternative methods

of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466-491.

George, D. y Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update*. Boston: Allyn & Bacon.

Gibbs, G. (2006). How assessment frames student learning. En C. Bryan y K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 23-36). London: Routledge.

Gibbs, G. y Dunbar-Goddet, H. (2007). *The effects of programme assessment environments on student learning*. York: Higher Education Academy.

Gibbs, G. y Simpson, C. (2003, Septiembre). *Measuring the response of students to assessment: The Assessment Experience Questionnaire*. Comunicación presentada en el 11th International Improving Students Learning Symposium. Hinckley.

Gibbs, G. y Simpson, C. (2004). Conditions under which assessment supports student learning. *Learning and Teaching in Higher Education*, *1*, 3-31.

Hattie, J. (1987). Identifying the salient facets of a model of student learning: A synthesis of meta-analyses. *International Journal of Educational Research 11*, 187-212.

Hattie, J. y Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77(1), 81-112.

Hambleton, R. K. (1996). Adaptación de test para su uso en diferentes idiomas y culturas: fuentes de error, posibles soluciones y directrices prácticas. En J. Muñiz (Ed.), *Psicometría* (pp. 67-89). Madrid: Universitas.

Lynn, M. (1986). Determination and quantification of content validity. *Nursing Research*, *35*, 382-85.

Ibarra, M. S. y Rodríguez, G. (2010). Aproximación al discurso dominante de la evaluación del aprendizaje en la universidad. *Revista de Educación*, *344*, 25-375.

Marsh, H., Martin, A. y Jackson, S. (2010). Introducing a short version of the Physical Self-Description Questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport and Exercise Psychology*, *32*, 438–482.

Martín-Albo, J., Núñez, J. L. y Navarro, J. G. (2009). Validation of the Spanish version of the Situational Motivation Scale (EMSI) in the educational context. *The Spanish Journal of Psychology*, *12*(2), 799-807.

Mateo, J. (2006). Claves para el diseño de un nuevo marco conceptual para la medición y evaluación educativas. *Revista de Investigación Educativa*, *24*(1), 163-186.

McDowell, L., Wakelin, D., Montgomery, C. y King S. (2011). Does assessment for learning make a difference? The development of a questionnaire to explore the student response? *Assessment & Evaluation in Higher Education*, *36*(7), 749-765.

Núñez, J. L., Martín-Albo, J. y Navarro, J. G. (2005). Validación de la versión española de la Èchelle de Motivation en Èducation. *Psicothema*, *17*(2), 344-349.

Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuk, C. N., Lohr, J. M., y Elwood, L. S. (2007). The disgust scale: Item analysis, factor structure, and suggestions for refinement. *Psychological Assessment*, *19*(3), 281-297.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education*, *16*, 129-150.

Reyes, C., Sosa, F., Marchena, R. y Marchena, J. (2013). Sistema de evaluación en las prácticas clínicas de la titulación de Medicina. *Revista de Educación Médica*, *16*(1), 7-11.

Richardson, J. (2005). Instruments for obtaining student feedback: A review of a literature. *Assessment & Evaluation in Higher Education*, *30*(4), 387-415.

Turner, G. y Gibbs, G. (2010). Are assessment environments gendered? An analysis of the learning responses of male and female students to different. *Assessment & Evaluation in Higher Education*, *35*(6), 687-698.